

RESEARCH ARTICLE

Two genetic loci associated with ankle injury

Stuart K. Kim^{1*}, John P. Kleimeyer², Marwa A. Ahmed³, Andrew L. Avins⁴, Michael Fredericson², Jason L. Dragoo², John P. A. Ioannidis^{5,6,7}

1 Dept. Developmental Biology, Stanford University Medical Center, Stanford, CA, United States of America, **2** Dept. Orthopaedic Surgery, Stanford University Medical Center, Stanford, CA, United States of America, **3** Dept. Physical Medicine and Rehabilitation, Harvard Medical School, Boston, MA, United States of America, **4** Kaiser Permanente Northern California, Division of Research, Oakland, CA, United States of America, **5** Dept. of Medicine, Stanford Prevention Research Center, Stanford University School of Medicine, Stanford, CA, United States of America, **6** Dept. of Health Research and Policy, Division of Epidemiology, Stanford University School of Medicine, Stanford, CA, United States of America, **7** Dept. of Statistics, Stanford University School of Humanities and Sciences, Stanford, CA, United States of America

* stuartkm@stanford.edu



OPEN ACCESS

Citation: Kim SK, Kleimeyer JP, Ahmed MA, Avins AL, Fredericson M, Dragoo JL, et al. (2017) Two genetic loci associated with ankle injury. PLoS ONE 12(9): e0185355. <https://doi.org/10.1371/journal.pone.0185355>

Editor: Yong-Gang Yao, Kunming Institute of Zoology, Chinese Academy of Sciences, CHINA

Received: July 20, 2017

Accepted: September 11, 2017

Published: September 28, 2017

Copyright: © 2017 Kim et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: These third party data are available from NIH GRASP. The authors did not have any special access privileges and interested researchers can access the data at <https://grasp.nih.gov/FullResults.aspx> (Trait (s): Ankle injury).

Funding: This work was supported by grants from the NIH (5R01AG025941) to SKK. Participant enrollment, survey and sample collection for the RPGEH were supported by grants from the Robert Wood Johnson Foundation, the Ellison Medical Foundation, the Wayne and Gladys Valley

Abstract

Ankle injuries, including sprains, strains and other joint derangements and instability, are common, especially for athletes involved in indoor court or jumping sports. Identifying genetic loci associated with these ankle injuries could shed light on their etiologies. A genome-wide association screen was performed using publicly available data from the Research Program in Genes, Environment and Health (RPGEH) including 1,694 cases of ankle injury and 97,646 controls. An indel (chr21:47156779:D) that lies close to a collagen gene, *COL18A1*, showed an association with ankle injury at genome-wide significance ($p = 3.8 \times 10^{-8}$; OR = 1.99; 95% CI = 1.75–2.23). A second DNA variant (rs13286037 on chromosome 9) that lies within an intron of the transcription factor gene *NFIB* showed an association that was nearly genome-wide significant ($p = 5.1 \times 10^{-8}$; OR = 1.63; 95% CI = 1.46–1.80). The *ACTN3* R577X mutation was previously reported to show an association with acute ankle sprains, but did not show an association in this cohort. This study is the first genome-wide screen for ankle injury that yields insights regarding the genetic etiology of ankle injuries and provides DNA markers with the potential to inform athletes about their genetic risk for ankle injury.

Introduction

Ankle sprains and strains are the most common musculoskeletal injuries in athletes, especially indoor or court sports [1–3]. Low ankle sprains occur with injury to ankle ligaments, most commonly the lateral ligament complex. Ankle strains describe pathologic stretching or tearing of muscle or tendon. Other ankle joint derangements, such as instability, may occur with insufficiency of the soft tissue restraints of the ankle or malalignment. These injuries (collectively referred to as ankle injuries) often occur in athletic activity with axial loading of an inverted, plantar-flexed foot as the most common mechanism. Ankle sprains are more common in women than men, and in children or adolescents compared to adults [2,3].

Foundation, and Kaiser Permanente. Development of genotypic data was supported by grant RC2 AG036607 from the National Institutes of Health. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Little is known about genetic factors that affect risk for ankle injury. One study reported an association of the R577X mutation in the *actinin 3* (*ACTN3*) gene with acute ankle sprain from a study involving 142 cases of injury ($p = .011$) [4]. *ACTN3* encodes alpha-actinin skeletal muscle isoform 3, which is an actin-binding protein expressed in all skeletal muscles. In principle, genetic studies such as this have the potential to identify contributing factors to ankle injuries and to provide diagnostic markers informing individuals about their personal risk for injury.

In order to identify genetic factors that may provide insight about ankle injuries, we screened the entire genome for loci associated with these injuries. We identified individuals who had suffered an ankle injury from a cohort of 99,342 patients in the Research Program on Genes, Environment, and Health (RPGEH) of the Kaiser Permanente, Northern California (KPNC) health plan. A gene association analysis revealed one locus on chromosome 21 associated with ankle injury with genome-wide significance, and a second locus on chromosome 9 with an association just below genome-wide significance. We re-tested the *ACTN3* SNP for association with ankle injury in our cohort but did not see a significant association.

Methods

A genome-wide association screen was performed for ankle injury using data from the genotyped Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort of the Research Program in Genes, Environment and Health (RPGEH). The data generation and data analysis pipeline have been previously described [5]. A complete description of the cohort and study design can be found in dbGaP (Study Accession: phs000674.v1.p1).

Our analysis cohort ($n = 99,342$) includes 57,606 females, 41,670 males, and 66 individuals of uncertain sex (Table 1). Sex was determined previously based on heterozygosity of the X chromosome (dbGAP Study Accession: phs000674.v1.p1). Moreover, our analysis cohort is ethnically diverse, including 83,264 European-White (EUR); 8,560 Latino (LAT) and 7,518 East Asian (EAS) individuals based on ancestry principle components.

Participants were genotyped at over 650,000 SNPs [6]. Genotypes were then imputed using standard procedures with a cutoff of $R^2 > 0.3$ [7–9]. The quality of the imputed data was previously validated in Jorgenson et al., 2015 [10].

Table 1. Demographic factors of the GERA study population used in genome-wide association analyses of ankle injury.

	Cases ^a	Controls	Overall
Subjects (%)	1,696 (1.7%)	97,646 (98.3%)	99,342
Sex (%)^b			
Female	1,049 (1.8%)	56,557 (98.2%)	57,606
Male	645 (1.5%)	41,025 (98.5%)	41,670
Undetermined	1 (1.5%)	65 (98.5%)	66
Ancestry (%)^c			
European	1,421 (1.7%)	81,843 (98.3%)	83,264
Latin American	154 (1.8%)	8,406 (98.2%)	8,560
East Asian	121 (1.6%)	7,397 (98.4%)	7,518
Age^d	60.1 (±13.1)	62.9 (±13.7)	62.8 (±13.7)

^a Cases with ankle injury as defined by individuals with one or more qualifying ICD-9, ICD-10 or CPT-4 code in their EHR. For details, see [Methods](#).

^b Sex/gender as determined by an individual’s genetic data, reported as the number and percentage of total.

^c Race/ethnic groups as determined by PCA on an individual’s genetic data from the GERA cohort. Reported as the number and percentage of total for each respective group.

^d Age at subject enrollment in the GERA cohort, reported as mean age with standard deviation.

<https://doi.org/10.1371/journal.pone.0185355.t001>

Determination of genetic ancestry was performed by principal component analysis (PCA), as previously described [5,11]. These ancestry principal components were used in the GWAS to adjust for genetic ancestry.

Phenotype definition

Ankle injury cases were identified in the GERA cohort based on clinical diagnoses and surgical procedures captured in the KPNC electronic health record system. The electronic health record includes reported injuries over the entire lifetime of the patients, including those that occurred prior to enrollment in KPNC as well as those that occurred after the genotyping analysis was performed, if reported by the patient and recorded by the physician. International Classification of Disease, Ninth Revision (ICD-9), International Classification of Disease, Tenth Revision (ICD-10) and Common Procedure Terminology, Fourth Edition (CPT-4) codes, were used to identify cases of ankle injury (Table 2). Table 2 includes codes for: 1) ankle sprain, 2) ankle strain, 3) surgical repair for disrupted ankle ligament and 4) joint derangement of the ankle or foot. Ankle sprains and strains are not differentiated in the ICD-9 codes, while they are in the ICD-10 codes. Joint derangement includes joint instability.

Table 2. Ankle injury phenotypes classified by ICD and/or CPT codes.

ICD-9 code	Code Description ^a	N
845.01	Sprain/Strain Deltoid Ligament	38
845.02	Sprain/Strain Calcaneofibular Ligament	21
845.03	Sprain/Strain Tibiofibular Ligament, Distal	34
845.09	Other Sprains/Strains Of Ankle	789
718.87	Other Joint Derangement, Not Elsewhere Classified, Ankle And Foot	444
ICD-10 code		
S93.401A	Sprain Of Unspecified Ligament Of Right Ankle, Initial Encounter	190
S93.401D	Sprain Of Unspecified Ligament Of Right Ankle, Subsequent Encounter	44
S93.401S	Sprain Of Unspecified Ligament Of Right Ankle, Sequela	1
S93.402A	Sprain Of Unspecified Ligament Of Left Ankle, Initial Encounter	165
S93.402D	Sprain Of Unspecified Ligament Of Left Ankle, Subsequent Encounter	45
S93.402S	Sprain Of Unspecified Ligament Of Left Ankle, Sequela	1
S93.409A	Sprain Of Unspecified Ligament Of Unspecified Ankle, Initial Encounter	3
S93.411A	Sprain Of Calcaneofibular Ligament Of Right Ankle, Initial Encounter	2
S93.412A	Sprain Of Calcaneofibular Ligament Of Left Ankle, Initial Encounter	2
S93.421D	Sprain Of Deltoid Ligament Of Right Ankle, Subsequent Encounter	1
S93.422A	Sprain Of Deltoid Ligament Of Left Ankle, Initial Encounter	1
S93.422D	Sprain Of Deltoid Ligament Of Left Ankle, Subsequent Encounter	2
S93.422S	Sprain Of Deltoid Ligament Of Left Ankle, Sequela	1
S93.431A	Sprain Of Tibiofibular Ligament Of Right Ankle, Initial Encounter	1
S93.432D	Sprain Of Tibiofibular Ligament Of Left Ankle, Subsequent Encounter	1
S93.432S	Sprain Of Tibiofibular Ligament Of Left Ankle, Sequela	1
S93.492A	Sprain Of Other Ligament Of Left Ankle, Initial Encounter	4
S93.492D	Sprain Of Other Ligament Of Left Ankle, Subsequent Encounter	1
S93.492S	Sprain Of Other Ligament Of Left Ankle, Sequela	1
CPT		
27695	Repair, Primary, Disrupted Ligament, Ankle; Collateral	15

^a International Statistical Classification of Diseases and Related Health Problems (ICD-9 or ICD-10) and Current Procedural Terminology (CPT-4) codes extracted from KPNC electronic health records of GERA cohort subjects.

<https://doi.org/10.1371/journal.pone.0185355.t002>

Genome-wide association and meta-analysis

Genome-wide association analyses of the GERA cohort genomic data were conducted as previously described [5,11]. SNP associations were tested with ankle injury with a logistic regression model using allele counts for typed and imputed SNPs in an additive genetic model for each of the race/ethnic populations. The model was adjusted for genetic sex, age at enrollment into the RPGEH cohort, race/ethnicity using principal components, and variations in genotyping protocol [5,11]. We used 10 principal components for European (EUR), 6 for Latin American (LAT) and 3 for East Asian (EAS). The final number of SNPs that were analyzed was 8,795,348 for EUR; 9,153,118 for LAT and 8,055,053 for EAS. To account for inflation due to population stratification, the genomic control parameter (λ_{gc}) was calculated: EUR (1.008), LAT (1.008), EAS (1.048). Subsequently, p-values were adjusted for genomic control in each population. Results from each population were combined by inverse-variance, fixed-effects meta-analysis as previously described [5,11]. SNPs that did not contain data for EUR were removed, as EUR comprises more than 80% of the cohort. The final number of SNPs that was analyzed in the fixed-effects meta-analysis was 8,183,964. Power calculations were made using the software at http://csg.sph.umich.edu/abecasis/cats/gas_power_calculator/index.html [12].

We examined the level of heterogeneity as previously described [5,11].

To perform the sensitivity analysis, the total set of 1696 cases was split into a subset that are known to involve an ankle injury (1275 cases) and another subset (421 cases) that are either ankle or foot derangements as they were identified solely by ICD-9 code 718.87 (Other Joint Derangement, Not Elsewhere Classified, Ankle And Foot). Logistic regression was used to calculate the association of chr21:47156779:D and rs13286037 with each subset of cases. Because the number of cases in the subgroups was smaller than in the total, only six principal components were used for the EUR ancestry group. P-values were calculated from a fixed-effects meta-analysis.

Further bioinformatics investigation of the top genome-wide significant loci from the meta-analysis was conducted as previously described [5,11].

Summary statistics for all SNPs from the fixed-effects meta-analysis are available at NIH GRASP: <https://grasp.nhlbi.nih.gov/FullResults.aspx>.

Ethical considerations

This study analyzed stored data from RPGEH subjects who consented to genomic testing and use of their genomic data, as well as health data from the KPNC electronic health record, for future research studies. The health and genotype data for the subjects were de-identified. All study procedures were approved by the Institutional Review Board of the Kaiser Permanente Research Institute.

Results

Study population and genotype information

We performed a logistic regression for DNA variants associated with ankle injuries. Ankle injuries (which refers to ankle sprains, strains and other derangements) were identified by ICD-9, ICD-10 and CPT codes (Table 1). There were 1,696 cases and 97,646 controls in the GERA cohort. Overall, the period prevalence of ankle injury was 1.7%. Participation in sports was not included in the electronic health record, and hence we were not able to determine the incidence rate for the subset of the population who were athletes. Men showed a lower incidence of ankle injury than women that was statistically significant ($p = 1.3 \times 10^{-3}$; OR = .85; 95% CI = 0.77–0.94), consistent with previous results [2,3]. There was a small difference in the age

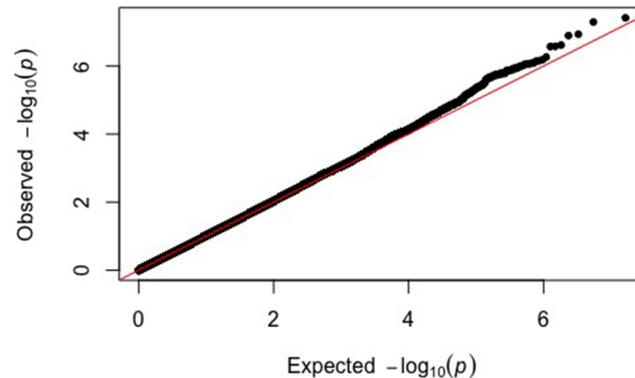


Fig 1. Quantile-quantile plot for genome-wide association analysis of ankle injury. The expected versus observed log transformed values for the 8,183,964 p-values from the meta-analysis are graphed. The y-axis shows the observed p-values and the x-axis shows the p-values expected by chance. The black dots represent the SNPs arranged by their observed p-values and the red line shows the expected trajectory if the SNPs had p-values expected by chance.

<https://doi.org/10.1371/journal.pone.0185355.g001>

of enrollment between the cases and controls (1.8 yrs) that was statistically significant ($p = 1.5 \times 10^{-8}$) (Table 1). One possibility is that this might be caused by an ascertainment bias where some elderly patients that enrolled in the RPGEH program might be systematically missed as cases if they incurred the ankle injury when they were young, before electronic records were in common practice.

Genome-wide study for association with ankle injury

The RPGEH cohort, genotyping data, methodological approach and logical flow presented here overlap those used in previous work by the same authors on MCL injury, shoulder dislocation, plantar fasciitis, ACL injury and Achilles tendon injury [5,11,13,14]. However, the analyses presented here present new results and concepts on the genetic basis for ankle injury, which has not previously been analyzed at the genome-wide level.

We compared the observed p-values to the distribution of p-values expected by chance in a Q-Q plot (Fig 1). The black dots deviate from the red line for the lowest observed p-values in the upper right hand corner, indicating that the observed association signals are significantly different than the signals that would be expected by chance.

The p-value for every SNP from the meta-analysis is shown in a Manhattan plot in Fig 2. Using $p = 5 \times 10^{-8}$ as a cut-off for genome-wide significance, chr21:47156779:D on chromosome 21 showed a genome-wide significant association with ankle injury ($p = 3.8 \times 10^{-8}$, Table 3). rs13286037 on chromosome 9 showed an association that was nearly genome-wide significant (5.1×10^{-8}) (Table 3). For both chr21:47156779:D and rs13286037, the minor allele frequency was below 5%, meaning that there were a relatively small number of individuals carrying either one or two copies of the risk allele. Fisher's exact test (i.e. linear discriminant analysis) is an alternative way to analyze the association of a DNA variant with ankle injury for small sample sizes. We repeated the analysis for association with ankle injury using Fisher's exact test for these two DNA variants, and obtained p-values that were similar although slightly less strong than the values using logistic regression. Specifically, the p-values using Fisher's exact test for chr21:47156779:D and rs13286037 were 4.8×10^{-8} and 9.5×10^{-8} , respectively.

Neither chr21:47156779:D nor rs13286037 were directly genotyped on the Affymetrix chips, but rather their genotype data was imputed (Table 3; S1 Table). For chr21:47156779:D, the R^2 value was 0.73, indicating that the genotype was only partially accurate using imputation

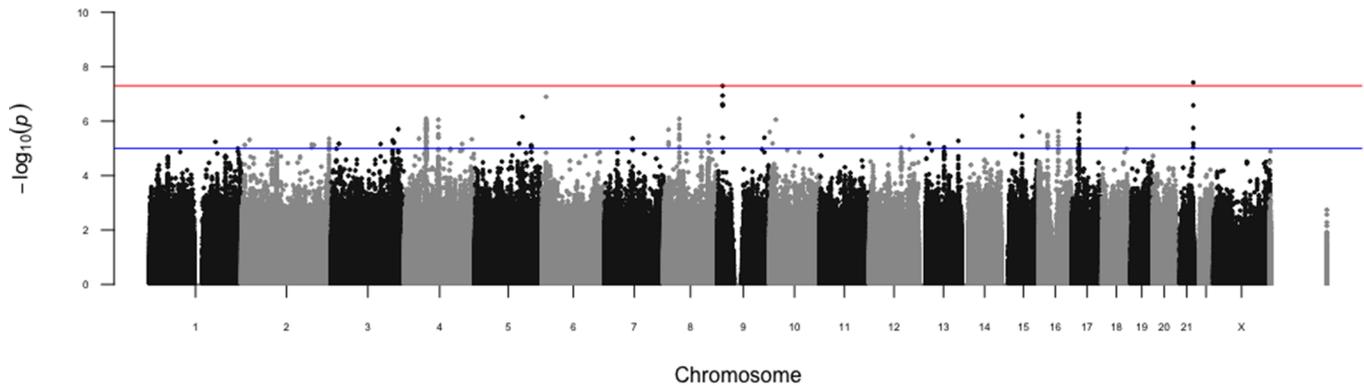


Fig 2. Manhattan plot for genome-wide association analysis of ankle injury. The $-\log_{10}$ p-values for association with ankle injury for SNPs from the meta-analysis are plotted by genomic position with chromosome number listed across the bottom. The y-axis shows the $-\log_{10}$ p-value for association with ankle injury. The blue line represents suggestive genome-wide significance ($p < 5 \times 10^{-5}$) and the red line represents genome-wide significance ($p < 5 \times 10^{-8}$).

<https://doi.org/10.1371/journal.pone.0185355.g002>

and thus care should be taken until true genotype data can be obtained. Inaccuracies in the genotype of chr21:47156779:D caused by imputation would not be expected to lead to a spurious association between chr21:47156779:D and ankle injury. Rather one might expect that the noise introduced by inaccurate imputation would weaken the association between the true genotype of chr21:47156779:D and ankle injury. For rs13286037, the R^2 value was 0.94 indicating that the imputed genotype is fairly accurate.

Of the 1696 cases of ankle injury, 421 were identified based solely on ICD-9 code 718.87, which pertains to instability or hypermobility of joints in either the ankle or foot. For this code, most of the injuries involve the ankle but some may involve foot derangements instead. Because some of the diagnoses with this code may have involved a foot rather than an ankle derangement, we wanted to know whether the association of chr21:47156779:D and rs13286037 was similar between the subset of cases known to affect the ankle (1275 cases) versus the subset identified by ICD-9 718.87 (421 cases). We split the cases into the two sub-groups, then repeated the logistic regression and meta-analysis for each sub-group (Table 4). The results indicate that both chr21:47156779:D and rs13286037 show an association with both the ankle sprain/strain and the ankle/foot derangement phenotypes. The odds ratios are similar for the two phenotypes. As expected, the p-values become weaker as the number of cases drops in each sub-group. These results indicate that the association of the top two SNPs with the ankle sprain/strain and the ankle/foot derangement phenotypes are qualitatively similar.

Table 3. Genome-wide association analyses for ankle injury.

Variant	Gene(s)	EA ^a	EAF ^b	P-value ^c	OR (95% CI) ^d
chr21:47156779:D ^e	<i>COL18A1</i>	A	.011	3.8×10^{-8}	1.99 (1.75–2.23)
	<i>SLC19A1</i>				
	<i>PCBP3</i>				
rs13286037 ^e	<i>NFIB</i>	A	.025	5.1×10^{-8}	1.63 (1.46–1.80)

^a Effect allele (minor allele).

^b Effect allele frequency in the control population.

^c P-value from fixed-effects meta-analysis.

^d Adjusted allelic odds ratio with 95% confidence interval.

^e Genotypes for these SNPs were imputed: chr21:47156779:D ($R^2 = 0.73$) and rs13286037 ($R^2 = 0.94$).

<https://doi.org/10.1371/journal.pone.0185355.t003>

Table 4. Sensitivity analysis for association with ankle sprains/strains and ankle/foot derangements.

chr21:47156779:D	Cases	P value ^a	OR (95% CI) ^b
All cases	1696	3.8x10 ⁻⁸	1.99 (1.75–2.23)
Ankle sprains/strains ^c	1275	1.2x10 ⁻⁵	1.89 (1.60–2.18)
Ankle/foot derangements ^d	421	5.2x10 ⁻⁴	2.26 (1.80–2.72)
rs13286037			
All cases	1696	5.1x10 ⁻⁸	1.63 (1.46–1.80)
Ankle sprains/strains ^c	1275	6.0x10 ⁻⁷	1.67 (1.47–1.87)
Ankle/foot derangements ^d	421	1.6x10 ⁻²	1.54 (1.19–1.89)

^a P-value from fixed-effects meta-analysis.

^b Adjusted allelic odds ratio with 95% confidence interval.

^c Cases known to an ankle sprain or strain.

^d Cases identified only by ICD-9 code 718.87 involving either an ankle or a foot derangement.

<https://doi.org/10.1371/journal.pone.0185355.t004>

Two loci associated with ankle injuries

For chr21:47156779:D, individuals that carried one copy of the risk allele (genotype A/AG) had a 1.86-fold increased risk of ankle injury compared to individuals with no risk alleles (genotype AG/AG)(Table 5). For rs13286037, individuals carrying one copy of the risk allele (A/T) had a 1.58-fold higher risk for ankle injury compared to individuals with no risk alleles (T/T)(Table 5). For both genetic variants, the risk was even higher in people carrying two copies of the risk allele, but there were too few such individuals for this result to be statistically significant (Table 5).

The GWAS results were analyzed to determine whether the association with ankle injury for either chr21:47156779:D or rs13286037 was stronger in some ancestry groups than in others, a phenomenon known as heterogeneity [15]. Table 6 shows the p-values and odds ratios for these two SNPs for each ancestry group. The logistic regression analysis did not converge on a p-value for the EAS ancestry for either chr21:47156779:D or rs13286037 due to limited number of cases. As expected, the smallest p-value for both SNPs was observed for the European population, since 82% of the cohort was European. For the LAT ancestry group, the p-values were 0.94 for chr21:47156779:D and 0.08 for rs13286037. The odds ratios for each race

Table 5. Genotype distributions for chr21:47156779:D and rs13286037.

chr21:47156779:D	A/A	A/AG	AG/AG
Cases	1	67	1,585
Controls	8	2,091	92,988
Overall	9	2,158	94,573
Risk for ankle injury	0.111	.0310	.0167
Relative risk for ankle injury ^a (95% CI)	6.65 (0.84–52.36)	1.86 (1.44–2.37)	1.00
rs13286037	A/A	A/T	T/T
Cases	3	131	1,612
Controls	83	4,845	95,394
Overall	86	4,976	97,006
Risk for ankle injury	.0361	.0263	.0166
Relative risk for ankle injury ^a (95% CI)	2.17 (0.69–6.89)	1.58 (1.35–1.90)	1.00

^a Risk relative to individuals homozygous for the protective allele (95% CI).

<https://doi.org/10.1371/journal.pone.0185355.t005>

Table 6. Association statistics for chr21:47156779:D and rs13286037 with ankle injury in individual ancestry groups.

Race	SNP	EA ^a	P-value ^b	OR	I ² ^d	Q ^e
				(95% CI) ^c		
EUR	chr21:47156779:D	A	1.57x10 ⁻⁸	2.07	42	0.19
				(1.61–2.66)		
LAT	chr21:47156779:D	A	0.94	1.04		
				(0.38–2.81)		
EAS	chr21:47156779:D	A	ND ^f	ND		
EUR	rs13286037	A	2.62x10 ⁻⁷	1.61	0	0.70
				(1.34–1.93)		
LAT	rs13286037	A	0.08	1.85		
				(.93–3.68)		
EAS	rs13286037	A	ND	ND		

^aEffect allele.

^bP value adjusted for lambda genomic inflation factor from fixed-effect meta-analysis.

^c Allelic odds ratio (95% confidence interval).

^d Percentage of variability between ancestry groups that is due to heterogeneity (95% confidence interval).

^e Cochran's Q, p-value that the association is different between ancestry groups.

^fNo data.

<https://doi.org/10.1371/journal.pone.0185355.t006>

were in the same direction and of similar magnitude. Using I² and Cochran's Q to assess heterogeneity, we saw no evidence of significant heterogeneity for rs13286037 between the EUR and LAT ancestry groups (Table 6). For chr21:47156779:D, the I² estimate was 42%, suggesting that there might be heterogeneity between the EUR and LAT ancestry groups. However, the 95% confidence interval for I² was 0–90, indicating that the presence and extent of heterogeneity is not certain.

chr21:47156779:D is located in the intergenic region between the protein-coding genes *COL18A1*, *SLC19A1* and *PCBP3* on chromosome 21 (Fig 3). *COL18A1* encodes the alpha chain of type XVIII collagen, which is a structural component of tendons and ligaments [16]. *SLC19A1* encodes Solute Carrier Family 19, which transports folate into cells [17]. *PCBP3* encodes Poly(rC)-Binding Protein 3, which binds poly(C) stretches in RNA [18]. rs13286037 is located in an intron of *NFIB*, which encodes a transcriptional repressor protein (Fig 4)[19].

We searched for a mechanism whereby these SNPs or a linked SNP might affect the activity of nearby genes to account for their effects on ankle injury. Using R²>0.7 as a threshold, there are three other SNPs (rs76694187, rs138382277 and rs118069956) that are in the same linkage disequilibrium block as chr21:47156779:D spanning about 71 kb on chromosome 21 (S1 Table). Any one of these four SNPs might be responsible for affecting risk for ankle injury, with the others showing an association simply due to linkage.

None of the four SNPs in the linkage block on chromosome 21 (chr21:47156779:D, rs76694187, rs138382277 and rs118069956) are in a coding region (Fig 3). ChIP seq experiments from ENCODE indicate that rs118069956 is located within a binding site for the transcription factors GATA2 and REST, raising the possibility that variation at rs118069956 might alter binding of these transcription factors and thereby influence expression of nearby genes [20]. rs118069956 is also located in a DNase I hypersensitive region, which is a region of open chromatin often caused by binding of transcription factors [20]. However, gene expression experiments have not yet been able to show that allelic variation in rs118069956 leads to changes in expression of nearby genes [21]. A second SNP in this region (rs138382277) might be responsible for variation in expression of a nearby long non-coding RNA. The Genotype-

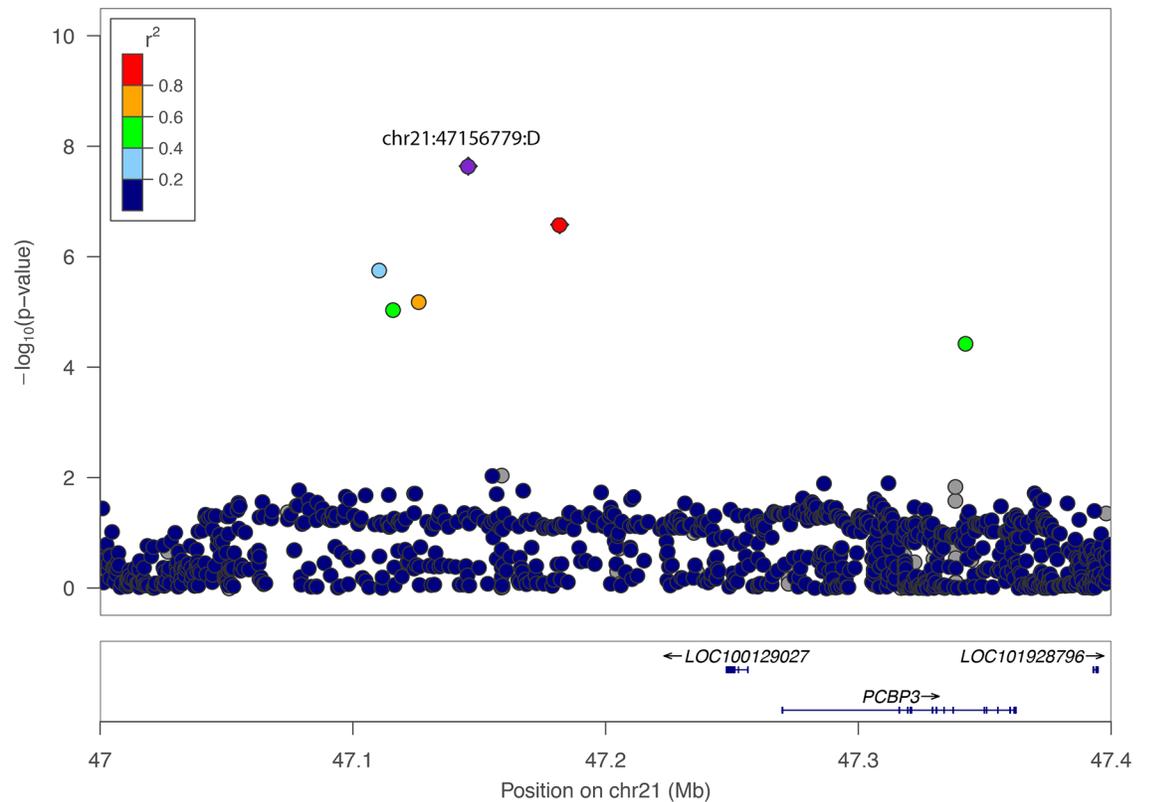


Fig 3. Regional-association plot for chr21:47156779:D with ankle injury. Tested SNPs are arranged by genomic position on chromosome 21 (x-axis) in a 600 kb window around the lead SNP chr21:47156779:D (purple diamond). The y-axis indicates $-\log_{10}$ p-values for association with ankle injury for each SNP. chr21:47156779:D is located in the intergenic region between *COL18A1/SLC19A1* and *PCBP3*. The location of *LINC00205* is not shown. The color of dots of the flanking SNPs indicates their linkage disequilibrium (R^2) with the lead SNP as indicated by the heat map color key.

<https://doi.org/10.1371/journal.pone.0185355.g003>

Tissue Exchange (GTEx) Portal has shown that rs138382277 is an expression quantitative trait locus (eQTL) for *LINC00205* (Long Intergenic Non-Coding RNA); specifically, the minor allele of rs138382277 is associated with lower expression of *LINC00205* and increased risk for ankle injury. *LINC00205* is located about 350 kb from rs138382277 with no known function, although this type of RNA has been found to alter chromatin structure and affect levels of gene expression [22].

For the locus on chromosome 9, rs35128680 is located 8.8 kb away from the sentinel SNP rs13286037 with $R^2 = 0.90$, indicating that the genotypes at these two SNPs are highly correlated (Fig 4). Data from the ENCODE projects show that rs35128680 is located within the central portion of the binding sites for three transcription factors (SMARCC1, TRIM28, MAX) and is also in a DNase I hypersensitive site [20]. These results suggest that rs35128680 might affect binding of one or more of these transcription factors, thereby changing expression of a nearby gene and affecting risk for ankle injury. However, expression data from the GTEx consortium has not yet shown that rs35128680 is an eQTL for *NFIB* or any other closely-linked gene, possibly because the relevant cell or tissue type for ankle injury has not been tested.

Re-testing *ACTN3* R577X for association with ankle injury

Shang et al. have reported that the R577X mutation in *ACTN3* (rs1815739) showed an association with acute ankle sprains [4]. We looked up the values for this SNP in our cohort and found that it did not show any signal of association ($p = 0.90$; OR = 1.00; 95% CI = 0.93–1.08).

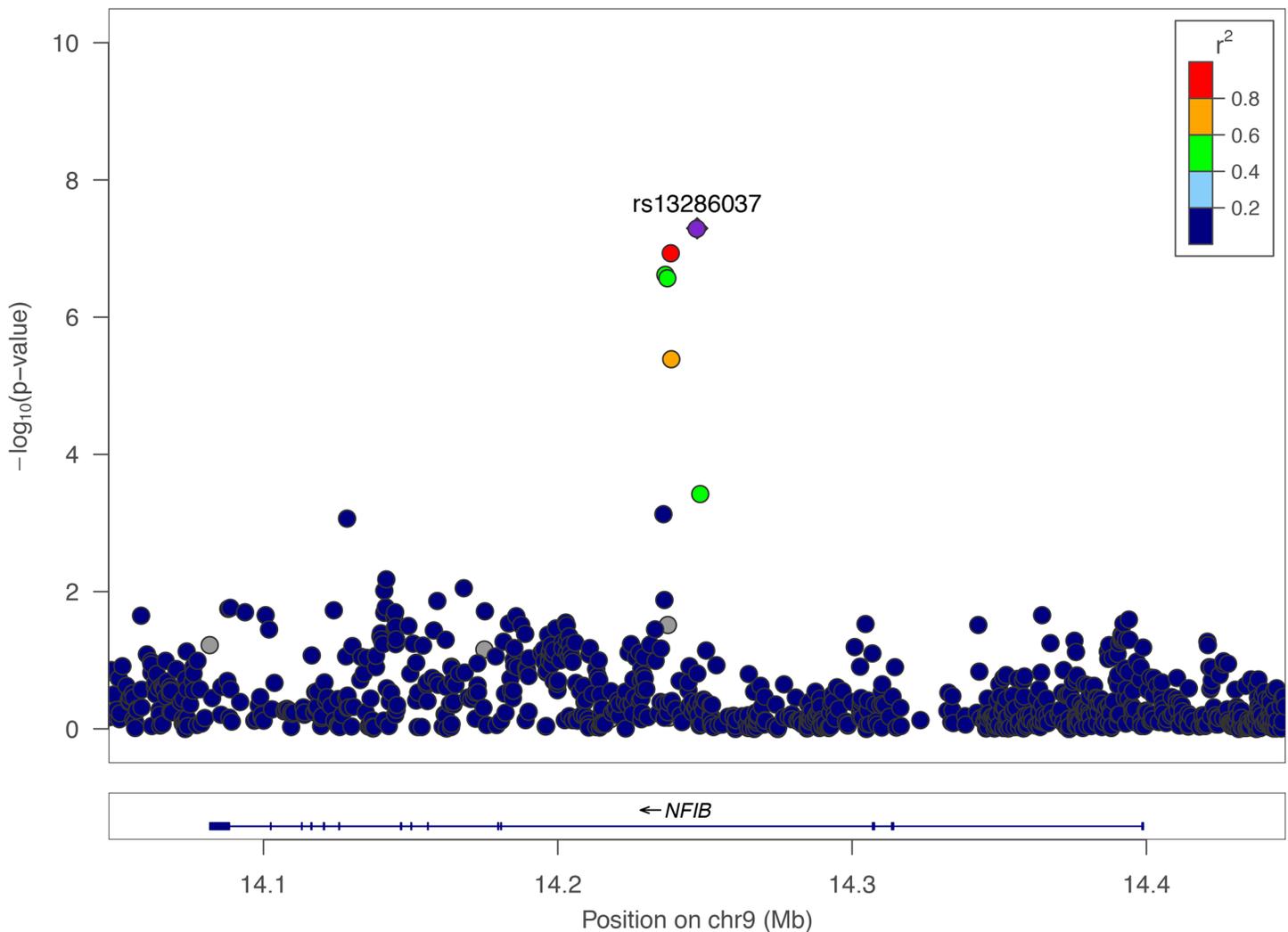


Fig 4. Regional-association plot for rs13286037 with ankle injury. Tested SNPs are arranged by genomic position on chromosome 9 (x-axis) in a 400 kb window around the lead SNP rs13286037 (purple diamond). The y-axis indicates $-\log_{10}$ p-values for association with ankle injury for each SNP. rs13286037 is located in the intron of *NFIB*. The color of dots of the flanking SNPs indicates their linkage disequilibrium (R^2) with the lead SNP as indicated by the heat map color key. Red dot indicates rs35128680, which is tightly linked to rs13286037 ($R^2 = 0.99$).

<https://doi.org/10.1371/journal.pone.0185355.g004>

Discussion

Ankle injuries, including sprains, strains, and other derangements, are common in athletes [2,3]. While ankle injuries are usually caused by an acute eversion or inversion, certain populations may be at greater risk of injury following such insults. Several risk factors have been described, however the genotype of the athlete may also impact an individual's risk for ankle injury, as well as the severity of injury and rate of recovery [2,3]. Such individuals may ameliorate their risk with preventative training, tailored conditioning and appropriate footwear [23–27].

Genetic markers for ankle injury

This study provides new information about the genetic factors associated for ankle injury. We demonstrated the first evidence for genetic factors affecting ankle injury with genome-wide significance, with large-scale genotype and phenotype data from the RPGEH with 99,342 individuals including 1,696 ankle injuries. Power calculations indicate that a cohort of this size

would have about a 90% chance of detecting a SNP with an association to ankle injury at genome-wide significance (assuming genotype relative risk of 1.7, minor allele frequency of 5%).

An indel (chr21:47156779:D) showed an association with ankle injury that was genome-wide significant (3.8×10^{-8}), and rs13286037 showed an association that was nearly genome-wide significant (5.1×10^{-8}). A previous study found that most genetic associations with moderate p-values (defined as $p \leq 10^{-7}$ and $p \geq 5 \times 10^{-8}$) were validated in subsequent studies [28]. These results should be replicated in an independent population in future studies. For chr21:47156779:D, we note that the imputed genotype was inferred with only 73% accuracy, indicating that the association with ankle injury should be viewed with some caution.

Potential genetic mechanisms for ankle injury

The chr21:47156779:D locus on chromosome 21 contains four linked variants that are located in the intergenic region between *COL18A1*, *SLC19A1* and *PCBP3* (Fig 3). *COL18A1* stands out as it encodes a collagen protein that might have a structural role in tendons or ligaments in the ankle [16]. chr21:47156779:D is an indel that has not been queried by either the ENCODE or GTEx project about whether it affects gene expression. rs118069956 is located 25 kb from chr21:47156779:D, and is situated in the binding region for the transcription factors GATA2 and REST [20]. A different SNP in this locus, rs138382277, is located 41 kb from chr21:47156779:D and is an eQTL for *LINC00205*, a long non-coding RNA situated 350 kb away. In summary, the association between genetic variation at this locus on chromosome 21 and ankle injury might involve changes in expression of a collagen gene or one of the two other nearby genes, or it might involve changes in expression of *LINC00205* located 350 kb away.

The locus on chromosome 9 contains two SNPs that show a moderate association with ankle injury (S1 Table). Both SNPs lie within an intron of a transcription factor gene *NFIB* (Fig 4). rs35128680 may affect expression of nearby genes as it is located within the binding regions of three transcription factors (SMARCC1, TRIM28, MAX)[20].

Predictive power of genetic testing for ankle injury

Individuals who have one copy of the risk allele for chr21:47156779:D (A) or rs13286037 (A) have an increased risk of 1.86-fold or 1.58-fold compared to individuals lacking a risk allele in our cohort, respectively. For the general population, a 58% or 86% increased relative risk for ankle injury may not warrant a change in lifestyle. For elite athletes participating in a jumping sport, however, this level of risk may warrant attention with regard to training regimen, because the consequences of injury can be substantial.

We were not able to replicate the association of the R577X mutation in *ACTN3* (rs1815739) with ankle injuries [4]. Power calculations show that our chance for replicating this result from a cohort of 1696 cases at $p \leq 0.05$ was 95% (assuming genotype relative risk ≥ 1.2). One explanation for the lack of validation is that the previous study looked at cases of acute ankle sprain in young, Chinese male soldiers. The difference in injury phenotype (ankle sprains/strains/derangements vs. acute ankle sprains), race (mostly European vs. Han Chinese) or population (general population in the Bay Area vs. soldiers) could account for the difference in the findings. Nevertheless, evidence from many other studies suggests that candidate gene associations need to be independently replicated, otherwise their credibility is low [29,30].

Limitations and future directions

As noted with previous analyses of this cohort, there are several limitations to this type of study [5,11]. First, the phenotypes were defined from codes contained in the electronic health records, which may be inaccurate. 877 of the 1696 cases (52%) were identified based on four

ICD-9 codes: 845.1, 845.2, 845.3 and 845.9. These four ICD-9 codes do not distinguish between ankle sprains and ankle strains. Furthermore, ICD-9 code 718.87 identified 421 cases of either ankle or foot derangements not elsewhere classified. Thus, some of the 718.87 diagnoses may have involved the foot rather than the ankle, or bony or paralytic derangements rather than sprains and strains.

Second, the ankle is composed of different ligaments, including the deltoid, lateral and tibio-fibular ligaments. It is unclear whether or not these ligaments have different aetiologies or underlying mechanisms for injury. The specific ligaments affected in each injury are usually not indicated by the ICD-9, ICD-10 or CPT codes in the electronic health record.

Third, the electronic health records do not distinguish between acute and chronic ankle injuries, which might have different aetiologies. Hence, the electronic health records do not provide knowledge about the nature or specific site of the ankle injury. The genetic association results presented here may derive from any one, or all, of the types of ankle injuries. Further investigation is warranted to study the differences in underlying genetic influences between specific ankle ligaments or chronic versus acute trauma on ankle injuries.

Fourth, the cohort included people regardless of whether or not they participated in a sport. We cannot document whether the statistical association of chr21:47156779:D and rs13286037 with ankle injury was derived predominantly from the subset of the population that were active in one or more sports.

Fifth, the number of individuals of Latin-American and East Asian ethnicity was relatively small (8,560 and 7,518, respectively). The association results for these ancestry groups are weaker than those from the European ancestry group, as would be expected due to smaller sample size. Heterogeneity analysis did not show a significant difference in the effect of either chr21:47156779:D or rs13286037 between the Latin-American or the East Asian ancestry group compared to the European ancestry group.

Sixth, the genotypes of chr21:47156779:D and rs13286037 were not directly measured but rather deduced by imputation. Care should be taken until the genetic association results can be replicated using direct genotyping of these loci in an independent cohort.

In the future, it will be important to replicate these gene association results with ankle injury in independent cohorts. Additional studies are warranted to begin to illuminate the underlying biological mechanism for the association of variation near *COL18A1*, *SLC19A1* and *PCBP3* on chromosome 21 and *NFIB* on chromosome 9 with ankle injury. It will also be interesting to perform the analysis on populations of athletes competing in sports with high rates of ankle injury, such as basketball or soccer. The results from these studies may reveal whether certain genetic polymorphisms such as chr21:47156779:D or rs13286037 could be used as diagnostic markers to help predict which athletes harbor a higher risk for ankle injury. Preventative measures could then be taken to alleviate that risk, thereby reducing the overall incidence of injury.

Supporting information

S1 Table. Linkage disequilibrium blocks on chromosomes 9 and 21.
(XLS)

Acknowledgments

The authors thank the Kaiser Permanente Northern California RPGEH team for access to data and assistance in data management. This work was supported by grants from the NIH (5RO1AG025941). Data for this study were provided by the Kaiser Permanente Research

Program on Genes, Environment and Health (RPGEH). Participant enrollment, survey and sample collection for the RPGEH were supported by grants from the Robert Wood Johnson Foundation, the Ellison Medical Foundation, the Wayne and Gladys Valley Foundation, and Kaiser Permanente. Development of genotypic data was supported by grant RC2 AG036607 from the National Institutes of Health. Information about data access can be obtained at: http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000674.v1.p1 and <https://researchbank.kaiserpermanente.org/for-researchers/>.

Author Contributions

Conceptualization: Stuart K. Kim.

Data curation: Stuart K. Kim.

Formal analysis: Stuart K. Kim, John P. A. Ioannidis.

Funding acquisition: Stuart K. Kim.

Investigation: Stuart K. Kim.

Methodology: Stuart K. Kim, John P. A. Ioannidis.

Resources: Stuart K. Kim.

Software: Stuart K. Kim.

Supervision: Stuart K. Kim.

Validation: Stuart K. Kim.

Writing – original draft: Stuart K. Kim.

Writing – review & editing: Stuart K. Kim, John P. Kleimeyer, Marwa A. Ahmed, Andrew L. Avins, Michael Fredericson, Jason L. Dragoo, John P. A. Ioannidis.

References

1. Shawen SB, Dworak T, Anderson RB (2016) Return to Play Following Ankle Sprain and Lateral Ligament Reconstruction. *Clin Sports Med* 35: 697–709. <https://doi.org/10.1016/j.csm.2016.05.012> PMID: 27543408
2. Fong DT, Hong Y, Chan LK, Yung PS, Chan KM (2007) A systematic review on ankle injury and ankle sprain in sports. *Sports Med* 37: 73–94. PMID: 17190537
3. Doherty C, Delahunt E, Caulfield B, Hertel J, Ryan J, Bleakley C (2014) The incidence and prevalence of ankle sprain injury: a systematic review and meta-analysis of prospective epidemiological studies. *Sports Med* 44: 123–140. <https://doi.org/10.1007/s40279-013-0102-5> PMID: 24105612
4. Shang X, Li Z, Cao X, Xie C, Gu M, Chen P, et al. (2015) The association between the ACTN3 R577X polymorphism and noncontact acute ankle sprains. *J Sports Sci* 33: 1775–1779. <https://doi.org/10.1080/02640414.2015.1012098> PMID: 25687200
5. Kim SK, Roos TR, Roos AK, Kleimeyer JP, Ahmed MA, Goodlin GT, et al. (2016) Genome-wide Association Screens for Achilles Tendon and ACL Tears and Tendinopathy *PLoS One* 12: e0170422.
6. Hoffmann TJ, Kvale MN, Hesselson SE, Zhan Y, Aquino C, Cao Y, et al. (2011) Next generation genome-wide association tool: design and coverage of a high-throughput European-optimized SNP array. *Genomics* 98: 79–89. <https://doi.org/10.1016/j.ygeno.2011.04.005> PMID: 21565264
7. Howie BN, Donnelly P, Marchini J (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 5: e1000529. <https://doi.org/10.1371/journal.pgen.1000529> PMID: 19543373
8. Howie B, Marchini J, Stephens M (2011) Genotype imputation with thousands of genomes. *G3 (Bethesda)* 1: 457–470.
9. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR (2012) Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet* 44: 955–959. <https://doi.org/10.1038/ng.2354> PMID: 22820512

10. Jorgenson E, Makki N, Shen L, Chen DC, Tian C, Eckalbar WL, et al. (2015) A genome-wide association study identifies four novel susceptibility loci underlying inguinal hernia. *Nat Commun* 6: 10130. <https://doi.org/10.1038/ncomms10130> PMID: 26686553
11. Roos AK, Avins AL, Ahmed MA, Kleimeyer JP, Roos TR, Fredericson M, et al. (2017) Two Genetic Loci Associated with Medial Collateral Ligament Injury. *Int J Sports Med* 38: 501–507. <https://doi.org/10.1055/s-0043-104853> PMID: 28482362
12. Skol AD, Scott LJ, Abecasis GR, Boehnke M (2006) Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat Genet* 38: 209–213. <https://doi.org/10.1038/ng1706> PMID: 16415888
13. Kim SK, Avins AL, Kleimeyer JP, Fredericson M, Ioannidis JP, Dragoo JL, et al. (2017) Genome-wide Association Study Reveals Two Genetic Variants Associated with Plantar Fasciitis or Fibromatosis. *Int J Sports Med* in press.
14. Kim SK, Kleimeyer JP, Ahmed MA, Avins AL, Fredericson M, Dragoo JL, et al. (2017) A Genetic Marker Associated with Shoulder Dislocation. *Int J Sports Med* 38: 508–514. <https://doi.org/10.1055/s-0043-106190> PMID: 28521375
15. Ioannidis JPA, Patsopoulos NA, Evangelou E (2007) Heterogeneity in meta-analyses of genome-wide association investigations. *PloS One* 2: e841. <https://doi.org/10.1371/journal.pone.0000841> PMID: 17786212
16. Oh SP, Warman ML, Seldin MF, Cheng SD, Knoll JH, Timmons S, et al. (1994) Cloning of cDNA and genomic DNA encoding human type XVIII collagen and localization of the alpha 1(XVIII) collagen gene to mouse chromosome 10 and human chromosome 21. *Genomics* 19: 494–499. <https://doi.org/10.1006/geno.1994.1098> PMID: 8188291
17. Matherly LH, Goldman DI (2003) Membrane transport of folates. *Vitam Horm* 66: 403–456. PMID: 12852262
18. Makeyev AV, Liebhaber SA (2000) Identification of two novel mammalian genes establishes a subfamily of KH-domain RNA-binding proteins. *Genomics* 67: 301–316. <https://doi.org/10.1006/geno.2000.6244> PMID: 10936052
19. Qian F, Kruse U, Lichter P, Sippel AE (1995) Chromosomal localization of the four genes (NFIA, B, C, and X) for the human transcription factor nuclear factor I by FISH. *Genomics* 28: 66–73. <https://doi.org/10.1006/geno.1995.1107> PMID: 7590749
20. Consortium EP (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* 489: 57–74. <https://doi.org/10.1038/nature11247> PMID: 22955616
21. Consortium GT (2015) Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* 348: 648–660. <https://doi.org/10.1126/science.1262110> PMID: 25954001
22. Rinn JL, Chang HY (2012) Genome regulation by long noncoding RNAs. *Annu Rev Biochem* 81: 145–166. <https://doi.org/10.1146/annurev-biochem-051410-092902> PMID: 22663078
23. Wedderkopp N, Kalltoft M, Lundgaard B, Rosendahl M, Froberg K (1999) Prevention of injuries in young female players in European team handball. A prospective intervention study. *Scand J Med Sci Sports* 9: 41–47. PMID: 9974196
24. Thacker SB, Stroup DF, Branche CM, Gilchrist J, Goodman RA, Weitman EA (1999) The prevention of ankle sprains in sports. A systematic review of the literature. *Am J Sports Med* 27: 753–760. <https://doi.org/10.1177/03635465990270061201> PMID: 10569362
25. Taylor JB, Ford KR, Nguyen AD, Terry LN, Hegedus EJ (2015) Prevention of Lower Extremity Injuries in Basketball: A Systematic Review and Meta-Analysis. *Sports Health* 7: 392–398. <https://doi.org/10.1177/1941738115593441> PMID: 26502412
26. McGuine TA, Brooks A, Hetzel S (2011) The effect of lace-up ankle braces on injury rates in high school basketball players. *Am J Sports Med* 39: 1840–1848. <https://doi.org/10.1177/0363546511406242> PMID: 21795671
27. Emery CA, Meeuwisse WH (2010) The effectiveness of a neuromuscular prevention strategy to reduce injuries in youth soccer: a cluster-randomised controlled trial. *Br J Sports Med* 44: 555–562. <https://doi.org/10.1136/bjsm.2010.074377> PMID: 20547668
28. Panagiotou OA, Ioannidis JPA, Genome-Wide Significance P (2012) What should the genome-wide significance threshold be? Empirical replication of borderline genetic associations. *Int J Epidemiol* 41: 273–286. <https://doi.org/10.1093/ije/dyr178> PMID: 22253303
29. Siontis KC, Patsopoulos NA, Ioannidis JP (2010) Replication of past candidate loci for common diseases and phenotypes in 100 genome-wide association studies. *Eur J Hum Genet* 18: 832–837. <https://doi.org/10.1038/ejhg.2010.26> PMID: 20234392
30. Ioannidis JPA, Tarone R, McLaughlin JK (2011) The false-positive to false-negative ratio in epidemiologic studies. *Epidemiology (Cambridge, Mass)* 22: 450–456.